Crystal structure analysis of c4763, a uropathogenic *Escherichia coli*-specific protein

Hun Kim,^{a,†} Jongkeun Choi,^{b,†} Doyoun Kim^{a,§} and Kyeong Kyu Kim^{a,*}

^aDepartment of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon 440-746, Republic of Korea, and ^bDepartment of Cosmetic Science, Chungwoon University, Hongseong Chungnam 350-701, Republic of Korea. *Correspondence e-mail: kyeongkyu@skku.edu

Received 23 June 2015

Accepted 7 July 2015

Edited by S. W. Suh, Seoul National University, Korea

† These authors contributed equally to this work.

§ Current address: Center for Synaptic Brain Dysfunctions, Institute for Basic Science, Daejeon 305-701, Republic of Korea.

Keywords: uropathogenic bacteria; crystal; allophanate hydrolase; urea; urinary-tract infection.

PDB reference: c4763, 5c5z

Supporting information: this article has supporting information at journals.iucr.org/f

Urinary-tract infections (UTIs), which are some of the most common infectious diseases in humans, can cause sepsis and death without proper treatment. Therefore, it is necessary to understand their pathogenicity for proper diagnosis and therapeutics. Uropathogenic *Escherichia coli*, the major causative agents of UTIs, contain several genes that are absent in nonpathogenic strains and are therefore considered to be relevant to UTI pathogenicity. c4763 is one of the uropathogenic *E. coli*-specific proteins, but its function is unknown. To investigate the function of c4763 and its possible role in UTI pathogenicity, its crystal structure was determined at a resolution of 1.45 Å by a multiple-wavelength anomalous diffraction method. c4763 is a homodimer with 129 residues in one subunit that contains a GGCT-like domain with five α -helices and seven β -strands. c4763 shows structural similarity to the C-terminal domain of allophanate hydrolase from *Kluyveromyces lactis*, which is involved in the degradation of urea. These results suggest that c4763 might be involved in the utilization of urea, which is necessary for bacterial survival in the urinary tract. Further biochemical and physiological investigation will elucidate its functional relevance in UTIs.

1. Introduction

Urinary-tract infections (UTIs), which are some of the most common infectious diseases in humans, are caused by the invasion of uropathogens through the urethra and their colonization of the bladder. In some cases, pathogens ascend *via* the ureter to the kidney and induce secondary infections. As a result, UTIs can be classified into lower UTIs, including cystitis and urethritis, and upper UTIs such as pyelonephritis. UTIs can be easily cured by antibiotics, but can sometimes result in sepsis or death if not treated properly. UTIs occur most commonly in women, and 30% of patients experience reoccurrence (Mobley & Warren, 1996). The major pathogen in UTIs is uropathogenic *Escherichia coli* (UPEC), which is responsible for more than 80% of cases of acute UTI (Braunwald *et al.*, 2001). The rate of antibiotic resistance of UPEC has increased to 20%, and the development of novel antibacterial therapeutics is therefore required. For these reasons, it is necessary to understand the bacterial pathogenesis of UPEC in detail.

Various virulence factors play a role in UPEC pathogenesis, such as bacterial adherence, toxin production, motility changes, metal acquisition and the evasion of host immune defences (Nielubowicz & Mobley, 2010). Comparative genomic hybridization (CGH) analyses of *E. coli* strains, including the most virulent UPEC strain CFT073, revealed 131 UPEC-specific genes that are not found in nonpathogenic strains (Lloyd *et al.*, 2007). Therefore, some of these genes are expected to be virulence factors and to be relevant to UTIs. From this point of

Table 1

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Peak	Inflection	Remote
Data collection			
Space group	$P2_12_12_1$		
Wavelength (Å)	0.97883	0.97923	0.96000
Unit-cell parameters			
a (Å)	57.61	57.63	57.63
b (Å)	66.34	66.38	66.36
c (Å)	69.84	69.89	69.87
$\alpha = \beta = \gamma$ (°)	90		
Resolution (Å)	30–1.45 (1.50–1.45)		
No. of observations	1394843 (137059)	697462 (68919)	697928 (67201)
No. of unique reflections	48282 (4759)	48327 (4753)	48241 (4766)
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
Multiplicity	28.9 (28.8)	14.4 (14.5)	14.5 (14.1)
Mean $I/\sigma(I)$	100.2 (14.0)	74.3 (9.3)	64.6 (7.0)
$R_{\text{merge}}^{\dagger}$ (%)	8.8 (38.9)	7.1 (37.1)	7.8 (44.6)
Refinement			
Resolution (Å)	30–1.45		
No. of reflections	48122		
$R_{\text{work}}^{\ddagger}/R_{\text{free}}$	0.189 (0.205)		
Molecules per asymmetric unit	2		
No. of atoms			
Protein	1983		
Water	228		
B factors (Å ²)			
Protein	15.7		
Water	22.5		
R.m.s.d. from ideal values			
Bond lengths (Å)	0.006		
Bond angles (°)	1.084		
Ramachandran plot (%)			
Most favoured regions	98.8		
Allowed regions	1.2		
Outliers	0.0		

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}, \quad \ddagger R_{\text{work}} = \frac{\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}$$

view, it is important to elucidate their structures and functions, which could provide new clues for understanding UPEC pathogenesis as well as identifying new drug targets. In this study, the structure of the c4763 protein, one of the UPEC-specific proteins, has been characterized and its functional implications are discussed.

2. Materials and methods

2.1. Protein expression and purification

The gene encoding c4763 was amplified by PCR from the genomic DNA of UPEC strain CFT073 and inserted into the pET-21a plasmid (Novagen, Massachusetts, USA) using NdeI and XhoI restriction sites. As a result, the recombinant c4763 contains an extra hexahistidine tag at the C-terminus. The resultant expression vector was introduced into *E. coli* B384 (DE3) competent cells. The transformed cells were then inoculated into M9 minimal medium supplemented with selenomethionine (SeMet) and other essential amino acids and cultured at 37°C. When the OD₆₀₀ of the cell culture reached 0.5, protein expression was induced by the addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and the

cells were cultured for a further 4 h. The cultured cells were harvested by centrifugation. The harvested cells were lysed by sonication in buffer *A* (20 mM Tris pH 7.5, 0.5 M NaCl, 20 mM imidazole, 1 mM PMSF, 5 mM DTT) and the soluble fraction was obtained by centrifugation. The supernatant was loaded onto a HiTrap nickel-chelating column (GE Healthcare, New Jersey, USA) and the bound protein was eluted using an imidazole gradient from buffer *A* to buffer *B* (20 mM Tris pH 7.5, 0.5 M NaCl, 1 M imidazole, 5 mM DTT). Fractions containing the c4763 protein were pooled, concentrated and subjected to a HiLoad Superdex 75 size-exclusion column (GE Healthcare, New Jersey, USA) pre-equilibrated with buffer *C* (20 mM Tris pH 7.5, 0.1 M NaCl, 5 mM DTT). The eluent was concentrated to 35 mg ml⁻¹ and frozen at -80°C until use.

2.2. Crystallization and data collection

Crystallization conditions were screened using SeMet-labelled c4763 at a concentration of 35 mg ml⁻¹ via the microbatch method under a 1:1 mixture of paraffin oil and silicone oil at 295 K. A drop consisting of 1 μ l screening solution and 1 μ l protein solution was placed into each well of a 72-well microplate. Diffraction-quality crystals were obtained from a crystallization solution composed of 0.1 M sodium/potassium phosphate pH 6.2, 0.2 M NaCl, 40% PEG 400 within a few days.

The crystal was immersed briefly into the cryoprotectant Paratone-N and flash-cooled in a nitrogen-gas stream at 100 K before X-ray exposure. The diffraction data were collected on the BL1A beamline at the Photon Factory, Tsukuba, Japan. For multiple-wavelength anomalous dispersion (MAD) phasing, data were obtained at wavelengths of 0.97883 Å (peak), 0.97923 Å (inflection) and 0.96000 Å (remote). Data processing and scaling were performed using the HKL-2000 suite (Otwinowski & Minor, 1997). Data-collection statistics are shown in Table 1.

2.3. Structure determination

The crystal structure of SeMet-labelled c4763 was determined by MAD phasing using the *AutoSol* program in the *PHENIX* package (Adams *et al.*, 2010) at 1.45 Å resolution. The initial model was constructed using the *AutoBuild* routine in *PHENIX*. Refinement and modelling were performed by *PHENIX* and *Coot* (Emsley & Cowtan, 2004), respectively. The final model was refined using diffraction data collected at a wavelength of 0.97883 Å. The stereochemical quality of the final model was evaluated by *MolProbity* (Chen *et al.*, 2010). Figures showing structures were prepared with *PyMOL* (DeLano, 2002) and the topology diagram was drawn using *Pro-origami* (Stivala *et al.*, 2011) and modified manually. The coordinates and structural factors of c4763 have been deposited in the Protein Data Bank with accession code 5c5z.

3. Results and discussion

The crystal structure of c4763 was determined at a resolution of 1.45 Å with an R_{work} and an R_{free} of 18.9 and 20.5%,

respectively. In the Ramachandran plot, 98.8 and 1.2% of the residues were assigned to the most favoured and the allowed regions, respectively. The crystal of c4763 belonged to space group $P2_12_12_1$. Two molecules were found in the asymmetric unit and were designated chains *A* and *B*. Consistent with these findings, purified c4763 eluted at a volume corresponding to a dimer in size-exclusion chromatography (data

not shown). Because of the weak electron densities for Met1 and Val129 of chain *A* and Thr128 and Val129 of chain *B*, the final model consists of residues 2–128 for chain *A* and residues 1–127 for chain *B*. The structural characterization in this study was performed using subunit *A* unless otherwise specified. The data-collection and refinement statistics are summarized in Table 1.

Figure 1

Crystal structure of the UPEC-specific protein c4763. (a) Ribbon diagram of the subunit structure of c4763, which consists of five α -helices and seven β -strands, which are coloured cyan and magenta, respectively. (b) Topology diagram of the subunit structure. The colour changes from blue to green to red indicate the direction of secondary structures from the N-terminus to the C-terminus. (c) Side and top view of two noncrystallographically related subunits. Twofold noncrystallographic symmetry is indicated by the line (left) and the oval (right).

Each subunit is composed of two domains. The N-terminal domain has an antiparallel β -sheet ($\beta 1$ – $\beta 2$ – $\beta 5$ – $\beta 6$ – $\beta 7$) packed by two long α -helices ($\alpha 1$ and $\alpha 2$) on one face, and the C-terminal domain contains a β -hairpin ($\beta 3$ – $\beta 4$) encompassed by three short α -helices ($\alpha 3$ – $\alpha 5$) (Figs. 1*a* and 1*b*). According

to a domain-architecture search using *CDART* (Geer *et al.*, 2002), c4763 contains a GGCT-like domain (also called an AIG2-like domain) and thus belongs to the GGCT-like superfamily of proteins. By comparison with proteins in the same structural family, it is expected that a ligand-binding

Figure 2

Structural comparison of c4763 and similar structures using ribbon diagrams. (*a*, *b*) Superposition of monomeric c4763 (cyan) with the C-terminal domain of the allophanate hydrolase monomer (blue) and the N-terminal domain of γ -glutamyl cyclotransferase (orange). (*c*, *d*) Structural overlap of the c4763 dimer with dimeric γ -glutamyl cyclotransferase and dimeric allophanate hydrolase. The same colour schemes are used as in (*a*) and (*b*). However, to clarify each subunit, the second subunits are shown in lighter colours.

Table 2

Hydrogen bonds in the intersubunit interface.

Residues involved in hydrogen bonds and the corresponding atoms are listed together with the hydrogen-bond distances.

Chain A	Distance (Å)	Chain B
Mse78 O	2.85	Lys88 N ^c
Glu80 O	2.78	Lys88 N ^c
Ile86 O	2.88	Ile86 N
Ile86 N	2.87	Ile86 O
Lys88 N ^c	2.69	Glu80 O
Lys88 N ^c	3.06	Mse78 O

pocket is present in the cleft formed between the two domains (Figs. 1*a* and 1*b*).

The two molecules in the asymmetric unit are related by twofold noncrystallographic symmetry (Fig. 1*c*). The pairwise root-mean-square deviation (r.m.s.d.) between the two molecules in the asymmetric unit is 0.7 Å for 126 C^α atoms. PISA analysis of c4763 reveals that the area of the monomer surface is 6764 Å² and that of the dimer surface is 11 428 Å² (Krissinel & Henrick, 2007), suggesting that 2100 Å² (accounting for 15% of each subunit surface) is buried by dimer formation. Intersubunit hydrogen bonds formed between residues located in the α2 helix, the β6 strand, a loop between α2 and β6 and a loop between the α3 and α4 helices (Table 2) predominantly contribute to the interaction between the two subunits. Hydrophobic residues also play a role in formation of the dimer, but neither disulfide bonds nor salt bridges are found in the subunit interface.

Since c4763 is classified as a hypothetical protein because of its low sequence similarity to other functionally identified proteins, its crystal structure was compared with known structures in order to obtain clues to its function and its contribution to UTIs. From structural alignment of c4763 using the DALI server (Holm & Rosenström, 2010), the two structures that showed the highest Z-scores were chosen for further analyses: the C-terminal domain of allophanate hydrolase (AH) from *Kluyveromyces lactis* (Fan *et al.*, 2013; PDB entry 4iss) and the N-terminal domain of human γ-glutamyl cyclotransferase (GGCT; Oakley *et al.*, 2008; PDB entry 2pn7). The C-terminal domain of AH shows the highest structural homology to c4763; they overlapped with an r.m.s.d. of 1.2 Å for 123 C^α atoms (Fig. 2*a*). However, c4763 is relatively less similar to the N-terminal domain of GGCT in its structural aspects, with an r.m.s.d. of 2.8 Å when 111 C^α atoms were superimposed (Fig. 2*b*). The sequence identity of c4763 to the C-terminal domain of AH and the N-terminal domain of GGCT is 27 and 15% for 123 and 111 residues, respectively.

Considering the area of the intersubunit interface of c4763 and its oligomeric state in solution, c4763 is highly likely to act as a dimer, and thus the dimeric c4763 was compared with homologous proteins. GGCT exists as a dimer consisting of two identical subunits. When the dimeric c4763 was superimposed onto the GGCT dimer, the second subunit (chain B) of c4763 overlaps poorly on chain B of GGCT, while both chain A of c4763 and GGCT show a good superposition (Fig. 2*c*). AH is also known to function as a dimer (Fan *et al.*, 2013).

Therefore, the c4763 dimer overlaps well with the C-terminal region of the AH dimer (Fig. 2*d*). Taking these results together, it is assumed that c4763 is more closely related to AH than to GGCT.

Based on the structural and sequential homology, it is expected that c4763 might be a functional homologue of AH. Allophanate hydrolase plays a crucial role in the urea-degradation pathway, together with urea carboxylase (UC). After UC carboxylates urea to give allophanate, AH hydrolyzes allophanate to ammonia and carbon dioxide in two steps: the N-terminal domain of AH converts allophanate to N-carboxycarbamate and the C-terminal domain of AH degrades N-carboxycarbamate to ammonia (Fan *et al.*, 2013). Since a limited number of *E. coli* strains are urease-positive (Mobley *et al.*, 1995) and no urease-encoding genes are found in the CFT073 strain, it is tempting to hypothesize that c4763 replaces part of the urease function. Therefore, it can be assumed that uropathogenic *E. coli* may employ c4763 as a fitness factor for its proliferation in the urinary tract by utilizing the abundant host urea as a nitrogen source. To verify the predicted function of c4763, we assayed the enzymatic activity of c4763 by measuring the amount of ammonia generated from allophanate, which is enzymatically linked to NADPH oxidation. However, we could not detect enzymatic activity of c4763 when allophanate was used as a substrate, possibly owing to the lack of an enzyme equivalent to the N-terminal domain of AH to initiate the hydrolysis of allophanate. Moreover, since N-carboxycarbamate, a reaction intermediate of AH, is unstable, we cannot use this compound as a substrate for enzymatic assay of c4763. Interestingly, in a BLAST search of the CFT073 genome using the N-terminal domain of AH as a query we found a protein in the CFT073 strain that is assigned as a putative amidase and has sequence homology to the N-terminal domain of AH. Therefore, it is tempting to hypothesize that c4763 might act together with the putative amidase in CFT073 in urea degradation. Further biochemical and structural studies will elucidate the biochemical and physiological functions of c4763.

Acknowledgements

This work was supported by the Agriculture, Food and Rural Affairs Research Center Support Program, Ministry of Agriculture, Food and Rural Affairs, a grant from the Next-Generation BioGreen 21 Program (SSAC PJ01107005), and grants from National Research Foundation of Korea (2011-0028878) (KK) and (NRF-2013R1A1A2059515) (KH).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Braunwald, E., Fauci, A. S., Kasper, D. L., Hauser, S. L., Longo, D. L. & Jameson, J. L. (2001). *Harrison's Principles of Internal Medicine*, 15th ed. New York: McGraw-Hill.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.

- Fan, C., Li, Z., Yin, H. & Xiang, S. (2013). *J. Biol. Chem.* **288**, 21422–21432.
- Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002). *Genome Res.* **12**, 1619–1623.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lloyd, A. L., Rasko, D. A. & Mobley, H. L. T. (2007). *J. Bacteriol.* **189**, 3532–3546.
- Mobley, H. L. T., Island, M. D. & Hausinger, R. P. (1995). *Microbiol. Rev.* **59**, 451–480.
- Mobley, H. L. T. & Warren, J. W. (1996). *Urinary Tract Infections: Molecular Pathogenesis and Clinical Management*. Washington: ASM Press.
- Nielubowicz, G. R. & Mobley, H. L. T. (2010). *Nature Rev. Urol.* **7**, 430–441.
- Oakley, A. J., Yamada, T., Liu, D., Coggan, M., Clark, A. G. & Board, P. G. (2008). *J. Biol. Chem.* **283**, 22031–22042.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. (2011). *Bioinformatics*, **27**, 3315–3316.